

JAN K. BRUECKNER

University of Illinois at Urbana-Champaign

Urban Sprawl: Lessons from Urban Economics

STRONG SENTIMENT AGAINST the phenomenon known as “urban sprawl” has emerged in the United States over the past few years. Critics of sprawl argue that urban expansion encroaches excessively on agricultural land, leading to a loss of amenity benefits from open space as well as the depletion of scarce farmland resources. The critics also argue that the long commutes generated by urban expansion create excessive traffic congestion and air pollution. In addition, growth at the urban fringe is thought to depress the incentive for redevelopment of land closer to city centers, leading to decay of downtown areas. Finally, some commentators claim that, by spreading people out, low-density suburban development reduces social interaction, weakening the bonds that underpin a healthy society.¹

To make their case, sprawl critics point to a sharp imbalance between urban spatial expansion and underlying population growth in U.S. cities. For example, the critics note that the spatial size of the Chicago metropolitan area grew by 46 percent between 1970 and 1990, while the area’s population grew by only 4 percent. In the Cleveland metropolitan area, spatial growth of 33 percent occurred over this period even though population *declined* by 8 percent.² Similar comparisons are possible for other cities.

This paper offers a technical development of arguments presented in a previous nontechnical paper (Brueckner, 2000b). I wish to thank Denise DiPasquale and the conference participants for helpful comments and suggestions.

1. For an excellent overview of the issues in the sprawl debate, see the twelve-article symposium published in the Fall 1998 issue of the *Brookings Review* (some of the articles are cited separately below).

2. See Nivola (1998, p. 18).

In response to concerns about sprawl, state and local governments have adopted policies designed to restrict the spatial expansion of cities. Twelve states have enacted growth management programs, with the best known being New Jersey's 1998 commitment to spend \$1 billion in sales tax revenue to purchase half of the state's remaining vacant land. Under a similar program, Maryland had allocated \$38 million to localities for purchase of nearly 20,000 acres of undeveloped land through 1998. Tennessee's 1998 antisprawl ordinance requires cities to impose growth boundaries or risk losing state infrastructure funds, mirroring an earlier, more stringent law in Oregon. Following the appearance of 240 antisprawl initiatives nationwide on November 1998 ballots, the November 2000 election saw many additional measures put before voters. Prominent statewide initiatives in Arizona and Colorado were defeated, but a number of local measures in California were approved.³

The stakes in the sprawl debate are substantial. Measures designed to attack urban sprawl would affect a key element of the American life-style: the consumption of large amounts of living space at affordable prices. Ultimately, an attack on urban sprawl would lead to denser cities containing smaller dwellings. The reason is that antisprawl policies would limit the supply of land for residential development, so that the price of housing, measured on a per-square-foot basis, would rise. In response to this price escalation, consumers would reduce their consumption of housing space, making new homes smaller than they would have been otherwise.

The goal of this paper is to assess the criticisms of urban sprawl and to identify appropriate remedies. To do so, it is important to start with a definition of sprawl. In this paper, urban sprawl will be defined as spatial growth of cities that is excessive relative to what is socially desirable. While no one doubts that spatial growth is needed to accommodate a population that is expanding and growing more affluent, the definition used here points to *excessive* growth as a problem.⁴

If the criticisms of sprawl are correct, then public policies should be altered to restrict the spatial expansion of cities. The resulting losses from lower housing consumption would be offset by gains such as improved access to open space and lower traffic congestion, and consumers on balance would be better

3. The information in this paragraph is drawn from Grant (2000); Richard Lacayo, "The Brawl over Sprawl," *Time*, March 22, 1999, pp. 45-46; Gurwitt (1999, pp. 19-20); and Rusk (1998, p. 14).

4. For an analysis of sprawl based on a much broader definition, see Downs (1999). Throughout the paper, the word "city" refers to an entire urban area. When necessary, a distinction is made between central city and suburbs.

off. But if the attack on sprawl is misguided, with few benefits arising from restricted city sizes, consumers would be worse off in the end. People would be packed into denser cities for no good reason, leading to a reduction in the American standard of living. The same conclusion would arise if some limitation of city sizes is desirable, but policymakers are overzealous. If only mild measures are needed to restrict urban growth that is slightly excessive, but draconian measures are used instead, consumers are likely to end up worse off.⁵

This paper identifies three fundamental forces that have led to the spatial expansion of cities: the growth of population; the rise in household incomes; and the decline in the cost of commuting. The paper identifies several reasons why the operation of these forces might be distorted, causing excessive spatial growth and justifying criticism of urban sprawl. These distortions arise from three particular market failures: the failure to account for the amenity value of open space around cities; the failure to account for the social costs of freeway congestion; and the failure to fully account for the infrastructure costs of new development. In each case, the market failure is analyzed, and an appropriate remedy is identified. The final pages discuss a common policy tool for dealing with sprawl, the urban growth boundary. Its pitfalls are noted, a few other issues discussed, and a conclusion offered.

The Fundamental Forces Underlying Urban Expansion

The fundamental forces underlying the spatial growth of cities are clearly delineated by the monocentric-city model.⁶ This model, which portrays the city as organized around a single, central workplace, can be criticized for failing to capture the recent evolution of U.S. cities, which now often contain multiple employment subcenters. However, since the monocentric model's main lessons about the spatial expansion of cities generalize to a more realis-

5. The ensuing discussion does not address a different issue frequently raised in criticisms of urban sprawl, namely, the proliferation of unattractive land uses such as strip malls and fast food outlets. Because this complaint concerns the character of development rather than its spatial extent, it lies outside the present definition of urban sprawl. Although ugly development cannot be banned, a remedy for this problem lies in the use of zoning regulations and other tools of urban planning, which allow land use to be channeled toward more aesthetic outcomes. These tools can complement the policies discussed below, which are designed to limit the extent, rather than the character, of development.

6. The model was developed by Alonso (1964); Mills (1967); and Muth (1969). See Mieszkowski and Mills (1993) for a nontechnical overview.

tic, polycentric urban area, it is acceptable to use the model in an analysis of urban sprawl.

In the model, urban residents commute to the central business district (CBD), where they earn a common income y . They incur a commuting cost of t per round trip mile, so that total commuting cost from a residence x miles from the CBD equals tx . Disposable income for households living at distance x is $y - tx$, and in the simplest version of the model, this income is spent on land, which represents housing, and a nonhousing composite good. These goods are denoted q and c , respectively.

To compensate for the long commutes of suburban residents, land rent per acre, denoted r , falls as distance to the CBD increases. This decline in rent in turn causes consumers to substitute in favor of land as distance increases, leading to greater land consumption (larger houses) in the suburbs. As explained in William C. Wheaton and later in Jan K. Brueckner, land rent and land consumption depend not only on distance x but also on income y and commuting cost t . In addition, r and q depend on the common utility level enjoyed by city residents, denoted u . These variables can thus be written $r(x, y, t, u)$ and $q(x, y, t, u)$.⁷

The urban utility level is endogenous and determined by the equilibrium conditions of the model. These conditions consist of two requirements: the city must fit its population; urban residents must outbid farmers for the land they occupy. Letting \bar{x} denote the distance to the edge of the city, r_a denote agricultural land rent, and n denote population, the well-known urban equilibrium conditions are

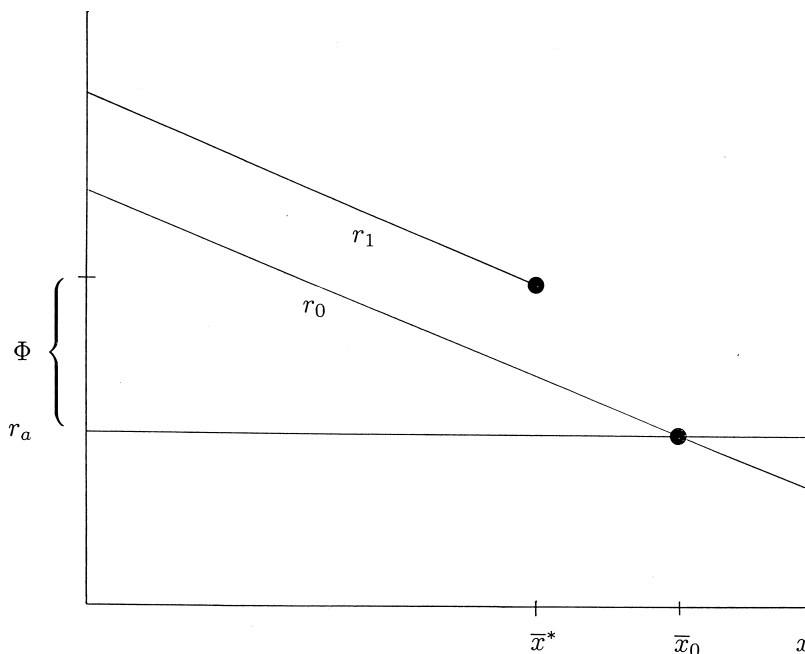
$$(1) \quad \int_0^{\bar{x}} \frac{2\pi x}{q(x, y, t, u)} dx = n$$

$$(2) \quad r(\bar{x}, y, t, u) = r_a.$$

To understand equation 1, note that since q is acres per person, $1/q$ gives people per acre, or population density. Multiplying density by the area of a ring of land ($2\pi x dx$) gives the number of people fitting in the ring, and integrating out to \bar{x} then yields the number of people fitting in the city, which must equal n . Condition 2 says that urban and agricultural land rents are equal at the edge of the city. Since r falls with x , this condition ensures that urban land rent is higher than r_a at all interior locations, as seen in figure 1 (the land rent curve r_o and the associated \bar{x} value of \bar{x}_o are relevant; the rest of the figure is dis-

7. Wheaton (1974); Brueckner (1987). See Brueckner (1987) for details.

Figure 1. Determination of City Boundary



cussed below). Together, conditions (1) and (2) determine equilibrium values for u and \bar{x} conditional on $n, y, t,$ and r_a .

The influence of these parameters on the city's spatial size can be derived by comparative-static analysis of (1) and (2), which was first presented by Wheaton.⁸ Wheaton established the following relationship:

$$(3) \quad \bar{x} = g[\underbrace{n, y}_{+}, \underbrace{t, r_a}_{-}].$$

Thus, the spatial size of the city grows as population n or income y increases, and falls as commuting cost t or agricultural rent r_a increases.

While the effect of n on \bar{x} is self-evident, the impacts of the other variables require more explanation. An increase in y affects the city's spatial size because urban residents demand more living space as they become richer. By itself, the greater demand for space causes the city to expand as housing consumption increases. But this effect is reinforced by the residents' desire to carry out their greater housing consumption in a location where housing is

8. Wheaton (1974).

cheap, namely, the suburbs. So the spatial expansion due to rising incomes is strengthened by an incentive for suburbanization.

A similar phenomenon occurs in response to investment in freeways and other transportation infrastructure. Because such investment makes travel faster and more convenient, thus reducing the cost of commuting, consumers can enjoy cheap housing in the suburbs while paying less of a commuting-cost penalty. As a result, suburban locations look increasingly attractive as commuting costs fall, which spurs suburbanization and leads to spatial growth of the city. In other words, \bar{x} rises when t falls.

Finally, equation 3 shows that, as agricultural land rent rises, competition from farmers for use of the land is more intense, and the city shrinks in response. Thus, the model predicts that in regions where agricultural land is productive and hence expensive, cities will be more spatially compact than in regions where land is unproductive and cheap. Productive agricultural land is therefore more resistant to urban expansion than unproductive land, reflecting the operation of Adam Smith's "invisible hand."⁹

Brueckner and David Fansler carry out an empirical test of the comparative-static predictions in (3).¹⁰ Using a 1970 sample of forty small to medium urbanized areas, they estimate a regression relating the city's land area to population, income, agricultural rent, and a commuting-cost proxy (the percentage of commuters using public transit). A high value for this proxy indicates a high t . Letting A denote land area, the regression results for a linear specification are as follows (absolute t statistics are in parentheses):

$$(4) \quad A = -41.07232 + .00041 * n - .03028 * r_a + .00620 * y - .24440 * t$$

(2.28) (10.03) (3.09) (3.03) (0.41)

Conforming to the predictions of the model, a city's land area rises as population and income increase, and land area falls with increases in agricultural

9. This effect counters a common claim among critics of urban sprawl, who sometimes argue that urban growth is an indiscriminate process, devouring agricultural land without regard to its worth. This view, however, is not consistent with the operation of a free market economy, where resources find their most productive uses. Concerns about loss of "scarce" farmland, often enunciated by critics of sprawl, are also misplaced. Because the value of farm output is fully reflected in the amount that agricultural users are willing to pay for the land, a successful bid by urban users means that society values the houses and other structures built on the land more than the farm output that is forgone. If farmland became truly scarce and in need of preservation, its selling price would be high, making the land resistant to urban encroachment.

10. Brueckner and Fansler (1983).

rent and commuting cost. The last effect is not statistically significant, however. The elasticity of A with respect to n is 1.1, indicating that a 1 percent increase in population leads to a 1.1 percent increase in land area. The elasticities for income and agricultural rent are 1.5 and -0.23, respectively.

The theory and evidence thus provide a compelling picture of how several fundamental forces (population and income growth combined with falling commuting costs) lead to urban spatial expansion. However, because the underlying model is monocentric, with all employment in the CBD, this explanation overlooks an important recent phenomenon in U.S. cities: job decentralization. It is natural to ask whether the movement of jobs out of central cities is an independent cause of spatial expansion or merely a consequence of the suburbanization of population.

Job relocation to the suburbs has been due in part to changes in the transport orientation of businesses. Rather than shipping their output through centrally located rail depots and port facilities, firms increasingly rely on truck transport and thus prefer the easy highway access (as well as the low cost) of suburban locations. However, the evidence shows that jobs also follow people.¹¹ In other words, job suburbanization is partly a response to the suburbanization of population, which occurs for the reasons discussed above. Thus, unlike the fundamental forces driving urban expansion, job suburbanization is partly an effect rather than a cause of this growth.

Given the confluence of an expanding national population, rising incomes, and falling commuting costs, it is not surprising that cities have expanded rapidly in recent decades. However, several market failures may distort the urban growth process, making this spatial expansion excessive relative to what is socially desirable.

These market failures, and potential remedies, are discussed in the next section.¹²

Sources of Market Failure in Urban Growth and Potential Remedies

Three market failures may lead to excessive spatial growth of cities. The first arises from a failure to take into account the social value of open space

11. Thurston and Yezer (1994) provide recent evidence as well as references to the prior literature on this topic.

12. It is well known that the urban equilibrium analyzed above, in which market failure is not present, is efficient. For inefficiency to arise, sources of market failure must be identified.

when land is converted to urban use. The second arises from a failure on the part of individual commuters to recognize the social costs of congestion created by their use of the road network, which leads to excessive commuting and cities that are too large. The third market failure arises from the failure of real estate developers to take into account all of the public infrastructure costs generated by their projects. This makes development appear artificially cheap from the developer's point of view, encouraging excessive urban growth.

Failure to Account for the Social Value of Open Space

To analyze the effect of this market failure, suppose that urban residents value the open space around the city. To simplify the discussion, let consumer preferences be given by the utility function $v(c, q, s)$, where s represents open space. Furthermore, let open space be measured by $s = B - \pi \bar{x}^2$, where B is the land area of the region containing the city, and $\pi \bar{x}^2$ is the land area of the city itself. Thus, s measures the amount of the region's space that is not occupied by housing. Although a more complex formulation might be more realistic, this simple framework captures the aesthetic and recreational benefits from the presence of open space near the city in an acceptable way.

With this modification of the model, the social value of the vacant land around the city includes the agricultural rent it earns and the open-space benefits it generates. In this situation, it can be shown that the condition determining the socially optimal allocation of land to urban use is

$$(5) \quad r(\bar{x}) = r_a + \int_0^{\bar{x}} \frac{2\pi x}{q} \frac{v_s}{v_c} dx \equiv r_a + \Phi,$$

where the arguments of the land-rent function other than \bar{x} have been suppressed for simplicity (compare 2). The integral in equation 5, which is denoted Φ , represents the social value of an acre of open space. This social value equals the marginal rate of substitution between s and c , which gives the open-space benefits per person in terms of the numeraire good, weighted by population at distance x ($2\pi x/q$) and summed over all x values in the city. Equation 5 thus requires that urban land rent at the edge of the city equals the social value of vacant land, which in turn equals agricultural rent plus the open-space value Φ .

Figure 1 shows the determination of the socially optimal \bar{x} , denoted \bar{x}^* . At \bar{x}^* , urban land rent equals $r_a + \Phi$ as required by equation 5. Note that the required differential between the urban and agricultural rents can only be achieved if \bar{x} shrinks below the value \bar{x}_0 where the two rents are equalized. This shrinkage, by reducing the supply of urban land, leads to an escalation of

land rent, as can be seen in the higher position of the rent curve r_1 relative to the original curve r_o . While this rent escalation reduces consumer purchasing power, the urban utility level nevertheless rises because of the offsetting benefits from additional open space around the city.

Figure 1 thus establishes that, in the presence of open-space benefits, the city's equilibrium spatial size, as represented by \bar{x}_0 , is too large. The problem is that, since intangible open-space benefits do not constitute part of the income earned by the land when it is in agricultural use, the disappearance of these benefits does not show up as a dollar loss when the land is converted to urban use. The invisible hand thus ignores open-space benefits, causing too much land to be allocated to urban use and leading to excessive spatial growth of cities.

A simple form of government intervention can remedy this problem: charging a development tax on each acre of land converted from agricultural to urban use. The development tax per unit of land is set equal to Φ (evaluated at the social optimum). With such a tax, equation 5 rather than equation 2 is the equilibrium condition for decentralized conversion of land, so that the outcome is optimal.

The difficulty, of course, is that implementing such a policy requires assigning a dollar value to the open-space benefits provided by an acre of land. Although economists have tried to estimate such values, the results are not sufficiently credible to use as a reliable basis for policy.¹³ This puts the policymaker in the position of having to guess the correct magnitude for a development tax. Because of the shortage of quantitative evidence on amenity

13. See Blomquist and Whitehead (1995) for an overview of contingent-valuation methods for measuring the value of environmental amenities. These methods rely on surveys that ask respondents for their willingness to pay (WTP) to preserve environmental amenities. For studies focusing specifically on the amenity value of vacant land, see Lopez, Shah, and Attolbello (1994) and Breffle, Morey, and Lodder (1998). Using previous contingent-valuation estimates, Lopez and others estimate the marginal amenity value of agricultural land for three different cases. For two cities in Massachusetts, the marginal values are \$8.80 and \$67.00 per acre, respectively, while the marginal value is \$31.00 for a city in Alaska. These numbers are expressed on an annualized basis, with present values approximately twenty times as large. Combining these open-space values with estimates of the demand for urban land, the authors conclude that, under an optimal allocation, between 3 and 20 percent more land would be allocated to agriculture across the three cases. Although key steps in the methodology are not clear, the resulting magnitudes are plausible. Less plausible values emerge from the more recent contingent-valuation study of Breffle, Morey, and Lodder (1998). The authors surveyed residents of Boulder, Colorado, asking their WTP to keep a 5.5-acre tract of land on the urban fringe undeveloped. Total WTP for residents within a one-mile radius of the tract was \$764,000, a number that exceeded the developer's \$600,000 bid for the site. Such large amenity valuations, which would virtually prohibit development if used as a basis for policy, suggest that contingent-valuation methods should be applied with care.

benefits, any open-space policy is fraught with difficulties and potentially counterproductive.

A further point is that the above model may not be accurate as a picture of how open-space benefits are generated. Rather than caring in the abstract about open space outside the city, consumers may be more affected by the availability of space in their immediate neighborhoods, in the form of city parks. One might argue that if city park land is adequately provided, the amount of open space outside of cities would not be a pressing concern of most urban residents. This, in turn, would undermine the case for a development tax like the one described above.¹⁴

Failure to Account for the Social Costs of Freeway Congestion

The second market failure that might affect the spatial sizes of cities arises through the activity of commuting. To understand this market failure, note first that commuting costs incurred by urban residents include the out-of-pocket costs of vehicle operation as well as the “time cost” of commuting. The latter cost measures the dollar value to the commuter of the time consumed while in transit. Together, these out-of-pocket and time costs represent the “private cost” of commuting, the cost that the commuter himself bears.

When the commuter drives on congested roadways to get to work, another cost is generated by his trip, above and beyond the private cost. This cost is due to the extra congestion caused by the commuter’s presence on the road. Thus, on congested roads, the true social cost of commuting for an individual includes the costs imposed on other commuters by the extra congestion that he creates. Note that while this extra congestion is slight, its impact is significant because many other commuters are affected.

Since these congestion costs are borne by others, the commuter himself has no incentive to take them into account. This missing incentive constitutes a market failure, and it means that commuting on congested roadways looks artificially cheap to individual commuters. The result is that congested roads are overused from society’s point of view.

To correct this problem, reducing road usage to socially optimal levels, several steps are appropriate. Some traffic should be diverted to off-peak hours, when roads are less congested, and some car commuters should switch to public transit. In addition, because of the overlooked social costs of commuting, the average commute distance is too long from society’s point of view, and it

14. Note that provision of more open space *within* cities would lead to an expansion rather than a contraction in their spatial sizes.

should be shortened. But an excessively long average commute means that cities are too spread out. Therefore, by causing people to commute too far, the market failure associated with freeway congestion can lead indirectly to urban sprawl.

Because the source of the problem is the individual's false perception of the costs of commuting, the remedy is to raise commuting costs by imposing a "congestion toll." Such a toll charges each commuter for the congestion damage he imposes on others. When a toll is levied, the out-of-pocket cost of rush hour commuting rises, and individuals have an incentive to shorten their commutes. Since this means living closer to one's job location, the ultimate effect is a spatial shrinkage of the city.

To see these arguments formally, let $T(x)$ denote the cost per mile of commuting at distance x from the CBD. This is simply the cost of crossing the narrow ring of land at x . Without congestion, $T(x)$ is simply equal to the exogenous constant t . But in the presence of congestion, $T(x)$ is given by

$$(6) \quad T(x) = t + f(P(x), R(x)),$$

where f is a function that captures congestion costs. Its second argument, $R(x)$, gives the amount of land at distance x devoted to roads. Since greater road capacity reduces congestion, f is decreasing in R . The first argument, $P(x)$, represents the traffic flow at x . Since the city is monocentric, $P(x)$ equals the number of people *living outside* x , which is written $\int_x^{\bar{x}} (2\pi z/q) dz$ (these people must cross the ring at x to reach the CBD). Since congestion worsens as traffic rises holding capacity constant, f is increasing in P . Note that while total commuting cost for a resident living at x is tx in the absence of congestion, this cost equals $\int_0^x T(z) dz$ with congestion.

Since an added commuter at distance x raises $P(x)$ by one, the commuter imposes extra congestion costs of $f_p(P(x), R(x))$ on each of the other commuters, where the subscript denotes partial derivative. Summing across commuters, the total congestion damage done by the extra commuter at distance x equals $P(x)f_p(P(x), R(x))$. To internalize the congestion externality, a congestion toll equal to this amount should be levied on each commuter passing through the ring at x . Note that since commuters symmetrically congest one another, each commuter pays this toll. Thus, the congestion toll at distance x is given by

$$(7) \quad \tau(x) = P(x)f_p(P(x), R(x)),$$

and for a commuter residing at x , the cumulative toll payment over his entire trip is $\int_0^x \tau(z) dz$.

Analysis of urban equilibria with congestion is difficult because commuting costs and the spatial distribution of population within the city are jointly determined as a result of the congestion phenomenon. In other words, where people live depends on commuting costs, but these costs in turn depend on where people live. Nevertheless, when the congestion externality is internalized via a toll, the effect on the spatial size of the city can be predicted intuitively. In particular, since the toll raises commuting costs, the impact on \bar{x} is similar to the effect of increasing t in a model without congestion. Equation 3 shows that \bar{x} falls as t rises, with higher commuting costs shrinking the city, and this same outcome occurs when congestion tolls are imposed in a model with congestion. Thus, in the absence of tolls, the city takes up too much space, as explained above.

The analysis establishing this outcome is clearest in Wheaton, although a large previous literature has investigated this type of model.¹⁵ Wheaton's numerical examples show that imposition of congestion tolls would reduce \bar{x} by 10 percent in his simulated city, from 28.9 miles to 25.9 miles. This result suggests that significant overexpansion of urban areas is caused by the failure to internalize the congestion externality.

Unlike the development tax discussed above, the proper magnitude of congestion tolls can be computed reliably, drawing on the wealth of accumulated knowledge about commuting behavior. A recent study by John Calfee and Clifford Winston, for example, computed the optimal toll as 27 cents per mile, which would generate roughly a \$6 charge for a 20-mile round-trip commute.¹⁶

Even though economists and transportation engineers uniformly endorse congestion tolls, they are seldom levied in practice. One problem is political: even though the revenue earned from tolls would allow other taxes to be reduced, commuters view tolls as a net tax increase, which creates opposition. Another problem is the daunting logistics of collecting tolls in a manner that does not impede traffic flows. In principle, technological advances can remove this obstacle by allowing toll charges to be tallied by electronic meters carried onboard autos. Low-tech solutions such as downtown parking taxes and costly bumper stickers that permit rush-hour usage of central roadways are also fea-

15. Wheaton (1998).

16. Calfee and Winston (1998). This estimate is computed using traffic volumes and capacity conditions from the ten largest metro areas, as well as the assumption that commuting time is valued at a traditional figure of 50 percent of the wage. If this valuation falls to 20 percent, as the authors argue is more realistic, the toll falls to 11 cents per mile. In addition, see Small (1992) for an extensive treatment of the theory and realities of congestion pricing.

sible. The latter approach was implemented in Singapore, while cities in Norway have experimented with more high-tech methods of collecting tolls.

By focusing just on the out-of-pocket and time costs of commuting, the preceding discussion ignores the resource costs of the transportation infrastructure used by commuters. Because such resource costs clearly constitute part of the social cost of commuting, a failure to make commuters pay for transportation infrastructure involves the same sort of underpricing of commute trips as the failure to levy congestion tolls. Although the gasoline tax functions as a user fee for the road network, with its revenues used for construction and maintenance, the prevailing tax levels in the United States are arguably far too low to cover these costs. As a result, commuters are undercharged for the resource costs of the infrastructure they use, which again encourages excessive commuting and urban sprawl.¹⁷

While an increase in the gas tax would remedy this second type of underpricing, a drawback is that the tax is paid regardless of the level of congestion encountered by the road user. Thus, unlike congestion tolls, the gas tax does not have the advantage of diverting traffic away from congested roads or congested travel times. Fortunately, this apparent dilemma in the choice between a user fee like the gas tax and congestion tolls is resolved by the theory of congestion pricing. The theory shows that the revenue from congestion tolls is likely to fully cover the infrastructure costs of the road network. More precisely, if roads are built with constant returns to scale and another natural technical assumption holds, then congestion-toll revenue exactly covers the cost of an optimal-size road.¹⁸

This result shows that if congestion tolls are levied, there may be no need to charge a separate user fee to pay for infrastructure costs. Although the gas tax therefore could be eliminated if a universal toll system were imposed, reliance on this tax is unlikely to end. Recognizing this likelihood, an increase in the gas tax would be one approach for attacking urban sprawl. While this approach is not ideal for the reasons discussed above, a higher gas tax would represent a practical means of remedying the underpricing of commute trips that contributes to sprawl.

Failure to Fully Account for the Infrastructure Costs of New Development

A third source of market failure that affects urban growth comes from the infrastructure costs generated by new development. When a new housing

17. See Mills (1999) for a clear presentation of this argument.

18. See Small (1992).

development is built, roads and sewers must be constructed, and facilities such as schools, parks, and recreation areas are needed. Homeowners, through the property tax system, pay for this infrastructure. The market failure arises because, under typical financing arrangements, the infrastructure-related tax burden on new homeowners is typically less than the actual infrastructure costs they generate. As a result, urban development appears artificially cheap, so that too much development occurs.

A formal analysis of this effect is presented by Brueckner.¹⁹ In his model, which is explicitly dynamic, a growing city invests in durable infrastructure in order to provide a constant level of public services to its residents. As population grows, the infrastructure stock must be enlarged to maintain the target level of services. Letting $n(T)$ denote the city's population at time T , the cost of the required infrastructure stock is $C(n(T))$.

When an additional resident is added to the city, requiring the conversion of one unit of land, the infrastructure stock must be enlarged, and the cost of doing so is given by the derivative $C_n(n(T)) > 0$. Because of the perfect durability of infrastructure, this cost of accommodating the new population is a one-time expense. With continuously growing population, however, a series of one-time costs must be incurred. Note that the annualized cost of the new infrastructure investment occurring at time T is $iC_n(n(T))$, where i is the interest rate.

From society's point of view, land is optimally converted to urban use when the net benefit from urban use of the land exceeds the agricultural rent r_a . This net benefit is equal to urban land rent minus the annualized cost of the infrastructure expansion needed to accommodate the additional population. Therefore, the condition for optimal conversion of the land is given by

$$(8) \quad r(T, \bar{x}(T)) - iC_n = r_a.$$

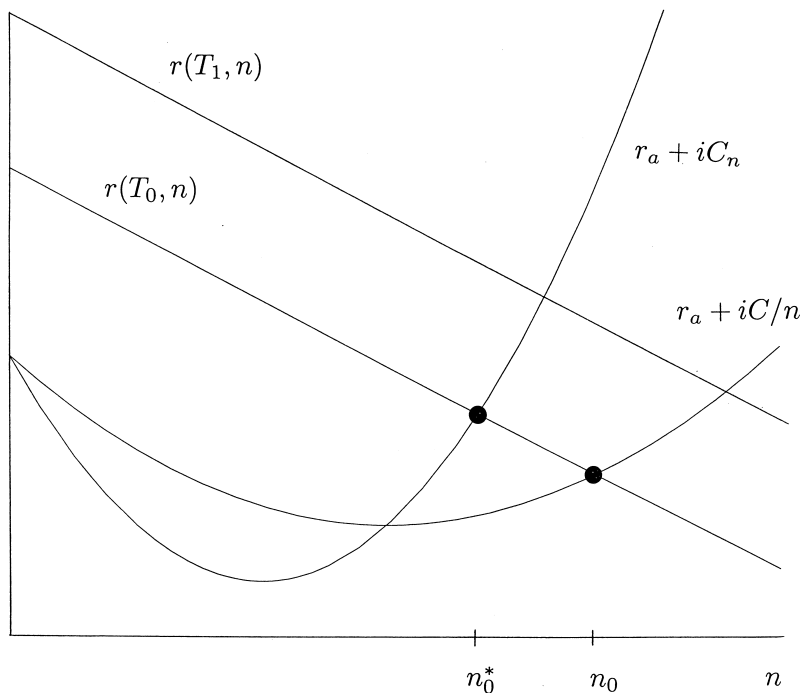
To emphasize the dynamic setting, time T appears as an argument of the land-rent function, with the function rising over time in exogenous fashion. In addition, note that $\bar{x}(T)$ gives the distance to the edge of the city at time T .

To provide a tractable analysis, Brueckner imposes several simplifying assumptions.²⁰ As noted above, land consumption per household is fixed at one unit, and the city is linear rather than circular, with a width also equal to unity. In this case, the urban population $n(T)$ and the boundary distance $\bar{x}(T)$ are the same. Then, substituting in (8) and rearranging, the equation can be written

19. Brueckner (1997a).

20. Brueckner (1997a).

Figure 2. Determination of City Populations



$$(9) \quad r(T, n(T)) = r_a + iC_n(n(T)),$$

where the argument of C_n is made explicit.

Equation 9 determines the socially optimal urban population at time T , and the solution can be seen graphically in figure 2. First note that the infrastructure cost function $C(\bullet)$ is assumed to generate U-shaped average and marginal cost curves, reflecting ranges of increasing and decreasing returns to scale in infrastructure provision. Then, the $n(T)$ that solves (9) lies where the curve corresponding to $r_a + iC_n$ (a translation of the marginal-cost curve) intersects the land-rent curve. Note that the latter curve, which gives rent at the edge of the city as a function of population, is downward sloping because rent falls with distance. The figure shows two rent curves corresponding to different times ($T_1 > T_0$) along with the optimal population size n_0^* at time T_0 .

This optimal solution can be contrasted with the one emerging from current institutions. For simplicity, Brueckner focuses on a decentralized system in which each landowner in the city shares equally in paying for the existing

stock of infrastructure.²¹ On the assumption that infrastructure is financed by infinite-maturity bonds, the required total payment at time T can be shown to equal $iC(n(T))$, the interest on the cost of the existing infrastructure stock. Assuming that this cost is spread evenly across all the developed land in the city, each acre of land incurs a cost of $iC(n(T))/n(T)$ at time T (recall that there are $n(T)$ acres of developed land in the city at this time). Note that this equal-payment assumption does not mimic a property tax regime, where higher-valued land close to the CBD incurs a greater tax liability. However, the assumption of equal payments serves as a convenient approximation.

With the tax burden on land equal to the average-cost expression $iC(n(T))/n(T)$, the condition governing decentralized conversion of land to urban use is

$$(10) \quad r(T, n(T)) = r_a + iC(n(T)) / n(T).$$

Note in equation 10 that the average-cost expression has replaced the marginal-cost term from equation 9. This difference yields a time T population for the city that differs from the socially optimal population, as can be seen in figure 2. The equilibrium population now lies at the intersection of the land-rent curve and the curve corresponding to $r_a + iC/n$, which is a translation of the average, rather than marginal, cost function.

Suppose the city has grown large enough to enter the range of decreasing returns to infrastructure provision, where the “average” curve $r_a + iC/n$ slopes up. Then, referring to figure 2, the equilibrium population at time T_0 , denoted n_0 , exceeds the socially optimal population n_0^* for that date. This relationship continues to hold as the land-rent curve shifts up over time, further enlarging the city’s population. The problem is that over this range, the social cost of adding infrastructure, given by iC_n , exceeds the average cost expression iC/n , which is what landowners face under equal payments. With the equal-payment regime understating the true cost of infrastructure, development appears artificially cheap, and too much of it occurs.

The reverse relationship holds when the city is still small, as can be seen in figure 2. When the equilibrium population is below the level that minimizes C/n , the socially optimal population is larger, rather than smaller, than the equilibrium population. Thus, insufficient urban expansion occurs when population lies in the range over which infrastructure exhibits increasing returns. However, since cities have expanded greatly as the U.S. population has grown, the range of decreasing returns may be relevant today.

21. Brueckner (1997a).

The remedy for the resulting urban overexpansion is to change the financing method for infrastructure. Rather than making all owners of developed land pay for the existing infrastructure as well as additions to the stock, a system of “impact fees” can be instituted. Under such a system, landowners whose land is converted at time T are charged a one-time fee of $C_n(n(T))$ to recover the infrastructure cost associated with the conversion. No future payments are required, with the cost of subsequent infrastructure additions paid for by landowners undertaking later development. Note that unlike marginal-cost charges in a static setting, impact fees fully pay for the stock of infrastructure. This follows because the fees exactly cover the cost of each increment to the stock as it is added.

In recent years, the use of impact fees has grown in many parts of the United States.²² Many communities in Illinois, for example, charge school impact fees, which defray the cost of new school construction. Properly computed, these fees may amount to nearly \$5,000 for a three-bedroom house.²³ In addition, road impact fees are often levied to defer the cost of expanding a city’s road network as population expands. Historically, impact fees have been challenged in the courts by real estate developers, who have contested the rights of communities to levy the fees or the methods used to calculate them. In some cases, the courts have ruled that impact fees do not properly reflect infrastructure costs, and they have promulgated standards to remedy such disparities.²⁴

Urban Growth Boundaries as a Remedy for Urban Sprawl

Three market failures leading to urban sprawl have been identified, and in each case, a specific remedy has been prescribed. These remedies (development taxes, congestion tolls, and impact fees) each involve use of the price mechanism to correct urban sprawl. Policymakers, however, often favor a much blunter instrument. This instrument is usually called an “urban growth boundary,” but other terminology is sometimes used. Rather than relying on

22. See Altshuler and Gomez-Ibanez (1993) for an overview of impact fee usage and Fischel (1990) for further discussion.

23. Calculations in Brueckner (1997b) compute school impact fees by combining data on school construction costs per square foot, space requirements per student, and the number of students generated by new houses of various sizes. For Illinois, the resulting fee for a three-bedroom house is \$4,560.

24. Altshuler and Gomez-Ibanez (1993).

taxes or congestion tolls to limit sprawl, an urban growth boundary (UGB) is a zoning tool that slows urban growth by banning development in designated areas on the urban fringe. In effect, imposition of such a boundary involves drawing a circle around a city and prohibiting development outside the circle.²⁵

A UGB is easy to implement, but it has great potential for misuse. The problem is similar to the one that arises in taxing development to preserve open space, namely, the need for guesswork. In particular, without a careful inquiry into the sources of market failure, policymakers cannot gauge the exact extent of urban overexpansion. As a result, there is a danger that a UGB may be much too stringent, needlessly restricting the size of the city and leading to an inappropriate escalation in housing costs and unwarranted increases in density.

For example, the failure to charge fully for the cost of infrastructure may result in a city that is 5 percent too large in area. Eager policymakers, however, may impose a growth boundary that ultimately makes the city area 15 percent smaller than in the absence of intervention. Such a draconian policy could be so harmful that society would be better off with no government intervention at all.

The way to avoid such errors is to attack urban sprawl at its source by imposing the specific remedies outlined above. Proper congestion tolls and impact fees can be computed with a high degree of reliability, ensuring that the resulting adjustments in urban spatial size are right from society's point of view. A development tax designed to preserve open space works equally well provided that a proper measure of open-space benefits can be computed.

The best known example of an urban growth boundary is from Portland, Oregon. Although some commentators claim that Portland's UGB is responsible for excessive house-price escalation in that city, as suggested by the above argument, others argue that the boundary is so loose that its price effects are negligible. This controversy illustrates an important point, namely, that there is no way to tell whether a UGB is set properly without focusing on the underlying market failures that lead to urban sprawl. Regardless of which view of the Portland case is correct, urban growth boundaries retain the potential for excessively restricting city sizes, and they should be used with great care.

A Numerical Example Showing the Effect of a Misplaced UGB

To get a sense of the effect of inappropriately restricting the spatial size of a city through an urban growth boundary, it is useful to consider a numerical

25. See Ding, Knaap and Hopkins (1999) for a recent analysis of UGBs.

example. To generate such an example, housing production is added to the simple model described earlier, following the approach of Brueckner.²⁶ Housing is produced with capital and land, and for purposes of the example, the production function is assumed to be of the Cobb-Douglas form with constant returns to scale. The utility function is also assumed to be Cobb-Douglas.

Parameter values for the simulated city are chosen realistically.²⁷ Then, two urban equilibrium conditions analogous to equations 1 and 2 are solved numerically using Mathematica. The results, which are presented in the first column of table 1, show that in equilibrium, the city has a radius of 30.8 miles and that its residents reach a utility level of 359.6. The price per unit of housing falls as distance from the CBD increases, leading to an increase in housing consumption over distance.²⁸ Combined with a decline in building heights over distance (not shown), this increase in dwelling size generates a dramatic decline in population density as distance increases. Finally, the table shows differential land rent in the city, which is the total land rent in excess of r_a generated between the CBD and \bar{x} . This differential rent constitutes the additional income that accrues to absentee landowners (over and above agricultural rent) because of the existence of the city.

To gauge the effect of an inappropriate UGB, suppose that the above equilibrium is not distorted in any way by market failure. The results of imposing a UGB in this situation can give a sense for the potential damage that can be done when an overly restrictive UGB is imposed under circumstances when market failures are present, but where only a mild restriction on \bar{x} is warranted. To this end, suppose that the UGB is set at fifteen miles, leading to

26. Brueckner (1987).

27. The simulated city is assumed to have a population of 800,000 households, representing 2 million people if households realistically contain 2.5 people. Household income is set at \$40,000 per year, and land rent per acre is set at \$250, which corresponds to a land value of \$5,000 per acre under a 5 percent discount rate (this yields rent of \$160,000 per square mile). Capital's exponent in the constant-returns production function is set at 0.75, and a multiplicative factor of 0.03 is applied to the production function so that realistic population densities are generated. The utility exponent on housing consumption is set at 0.5. Finally, the commuting cost parameter is set at \$500, a value that includes both money and time cost, as follows: Assuming an out-of-pocket cost of \$0.30 per mile and 250 round trips per year leads to a value of \$150 per year for the money cost of commuting per mile. A yearly income of \$40,000 implies an hourly wage of \$20, which yields a time cost per mile of \$0.66 assuming a traffic speed of 30 miles per hour (commuting time is valued at the full wage). Time cost is then \$330 per mile per year, yielding a total money plus time cost of \$480, which is rounded up to \$500.

28. Note that units of measurement for housing consumption (square feet, square meters, and so on), can be specified arbitrarily. A rescaling of consumption, however, would require an opposite rescaling of the price per unit.

Table 1. The Effects of an Urban Growth Boundary

<i>Item</i>	<i>Equilibrium city</i>	<i>City with UGB^a</i>
\bar{x}	30.8	15.0
Utility	359.6	335.5
<i>Population density</i>		
At CBD	52,098	90,914
At $x = 15$	12,178	21,252
At \bar{x}	1,734	...
<i>Housing price</i>		
At CBD	3,093	3,554
At $x = 15$	2,042	2,346
At \bar{x}	1,170	...
<i>Housing consumption</i>		
At CBD	6.47	5.63
At $x = 15$	7.96	6.92
At \bar{x}	10.51	...
Differential land rent	2.82×10^9	3.45×10^9

a. Income increase required to offset utility loss from UGB: \$2,942 (7 percent of original income).

more than a 50 percent reduction in the radius of the city. The consequences of this restriction are shown in the second column of table 1.

Utility falls from 359.6 to 335.5, so that the urban residents are worse off. To see the reasons for this loss, note that the price per unit of housing rises throughout the truncated city, yielding a decline in housing consumption. This dwelling-size reduction tends to raise population density, an effect that is strongly reinforced by an increase in building heights (not shown). As a result, the city's population density rises dramatically. Finally, differential land rent increases substantially, indicating that absentee landowners are better off under the UGB.

Thus, the results show that by packing people into a smaller city, the UGB raises the price of housing and cuts housing consumption. To get a sense of the magnitude of the resulting welfare loss, consider the following exercise. With the UGB in place, let consumer income be increased until the utility level achieved is the same as the equilibrium level from the first column of the table. The required increase in income, which gives a measure of the consumer welfare loss from the UGB, is equal to \$2,942. Thus, canceling the welfare loss requires a 7 percent increase in income.

Recognizing that the UGB imposes a very dramatic 50 percent reduction in the city's radius, this compensating income increase seems modest in size. This reflects two facts. First, population density in the region beyond a dis-

tance of fifteen miles is relatively low, so that the share of the population needing reaccommodation after imposition of the UGB is not commensurate with the land area reduction. Second, the displaced population is mostly housed through increases in building heights, without a dramatic decline in housing consumption.

While urban residents are hurt, an overall efficiency verdict on the UGB must also take into account the welfare of absentee landowners. To see that this overall verdict is negative, suppose that the government were to raise the income of each household so as to cancel the loss from the UGB.²⁹ Could the government recover the cost of the subsidy by taxing away the resulting increment in differential land rent? Calculations show that the land-rent increment falls well short of the required transfer to households.³⁰ As a result, the government cannot compensate consumers for the effect of the UGB while maintaining the incomes of absentee landowners, indicating that the UGB reduces overall welfare. Thus, the lesson of the analysis is that unwarranted restrictions in the spatial sizes of cities harm urban residents while lowering overall welfare.³¹

Although the model underlying the numerical example has a single type of household, the real-world impact of a UGB will be felt by consumers from different income groups. Moreover, it is likely that low-income households will be more adversely affected than the rich by any UGB-induced escalation of housing prices. Reflecting this possibility, consumer groups concerned about adverse effects on housing affordability joined housing developers in opposing several of the antisprawl measures appearing on the November 2000 ballot.³² Furthermore, if the environmental benefits that may result from an attack on sprawl constitute a luxury good, valued more by high- than low-income households, incidence of such policies may be skewed even more in favor of the well-off.

29. The required payment of \$2,942 to each of 800,000 households would involve a total transfer of \$2, 354 x 10⁶. The resulting higher incomes would in turn raise urban land rents beyond the pretransfer level, with differential land rent in the city rising to \$3.74 x 10⁹.

30. With differential rent prior to imposition of the UGB equal to \$2.82 x 10⁹, the increment equals \$920 x 10⁶, which falls well short of the required transfer of \$2,354 x 10⁶.

31. While there is little empirical evidence on the effects of UGBs in the United States, a number of empirical studies show the effects of other types of land-use restrictions on U.S. housing prices, with positive impacts typically found. For a survey of such studies, see Fischel (1990). UGBs are also found elsewhere in the world, and their harmful effects on housing affordability have been most thoroughly studied in the case of Korea. See Kim (1993) and Son and Kim (1998).

32. Habitat for Humanity joined housing developers in opposing antisprawl measures in Arizona and Colorado. See Richard A. Oppel Jr., "Efforts to Restrict Sprawl Find New Resistance from Advocates for Affordable Housing," *New York Times*, December 26, 2000, p. A18.

Developing this theme, a recent theoretical literature on urban growth controls abstracts from the market failures considered above and portrays growth-control policies as a way for landowners to raise their incomes (and perhaps their quality of life) at the expense of renters. The latter group pays higher housing prices as a result of the growth control's restriction on expansion of the city.³³

Other Factors Contributing to Sprawl

While market failures discussed earlier would appear to be prime culprits in generating excessive spatial growth of cities, additional forces contributing to this outcome can be identified. The first is another fiscal effect, which arises from the process of "voting with one's feet." This phrase refers to the tendency of high- and middle-income consumers to form separate jurisdictions for the provision of public goods such as education, public safety, and parks. Such jurisdictions tend to be created on the urban fringe, which exacerbates the tendency toward urban expansion.

As explained by Charles M. Tiebout, the goal of well-off consumers in forming such separate jurisdictions is to gain control over the level of public spending, which can then be set high enough to provide the high-quality schools and public services that such consumers demand.³⁴ An additional benefit comes from avoiding the need to subsidize the public consumption of poor households, who contribute little of the tax revenue required by local governments. To protect these benefits, the residents of suburban communities often impose minimum-lot-size restrictions and other fiscal zoning regulations designed to deter poor households from entering the community.

One way to diminish this tendency for Tiebout sorting (thus limiting the resulting urban expansion) is through a metropolitan taxing authority. Such an authority would divert funds from well-off suburban communities to the poor central city, limiting the gains from formation of such communities. However, political opposition from well-off households dooms most attempts to create metropolitan governments.³⁵

A number of other fiscal effects may contribute to urban sprawl. One such effect arises through the federal tax subsidy to owner-occupied housing, which

33. See, for example, Brueckner (1995); Helsley and Strange (1995).

34. Tiebout (1956).

35. See Orfield (1998) for an instructive discussion of the politics of metropolitan government in Minneapolis.

arises because imputed rental income is untaxed. Harvey Rosen shows that if imputed rent were instead taxed, housing consumption for homeowners would fall by 10 to 20 percent, with the exact number depending on household income.³⁶ Since the resulting reduction in dwelling sizes would reduce the consumption of land, the spatial sizes of cities would ultimately fall in the absence of the housing tax subsidy.

It is interesting to note, however, that another set of federal policies, those designed to maintain farm incomes, may offset the sprawl-inducing effects of the federal tax subsidy to homeowners. By raising the income-producing potential of the land, policies such as farm price supports tend to increase agricultural land rent, which has the effect of restricting, rather than encouraging, urban spatial expansion (recall equation 3). While these two policies may thus have offsetting effects on urban expansion, the policies in any case are designed to promote social goals (homeownership, the family farm) that are separate from the issue of urban sprawl. As a result, attempts to alter the policies to address the sprawl problem would probably be unwarranted.

Finally, another fiscal force arising through the property tax may also contribute to urban sprawl. As shown by Brueckner and Hyun-A Kim, the property tax reduces the intensity of land development (that is, building heights), which lowers population densities.³⁷ Lower densities, in turn, cause cities to spread out, creating sprawl. However, Brueckner and Kim's analysis identifies a countervailing effect that arises through the property tax's tendency to reduce dwelling sizes, which raises population densities. While the net effect of the tax on urban spatial size is thus ambiguous, simulation results suggest that it may be positive in a realistic model, making the property tax a potential culprit in the excessive spatial expansion of cities.

Byproducts of an Attack on Sprawl

An attack on urban sprawl might produce several byproducts. These include upgrading and redevelopment in central neighborhoods, which help to reverse the process of central-city decay. In addition, the higher densities generated by an attack on sprawl may improve the quality of urban life by fostering social interaction.

36. Rosen (1985).

37. Brueckner and Kim (2000).

As noted earlier, many commentators argue that excessive urban spatial growth contributes to the decay of central cities by reducing the incentive to redevelop land near the center. Central-city decay, however, would be a problem even in the absence of the market failures leading to urban sprawl. The reason is that the suburbanization forces generated by rising incomes and falling commuting costs reduce the demand for aging central-city housing, depressing its price and diminishing the incentive for upgrading and redevelopment. By inappropriately increasing the supply of developed land, overexpansion of cities exacerbates this tendency by putting further downward pressure on housing prices. Thus, the incentive for upgrading and redevelopment of aging dwellings is further reduced.

If, however, sprawl is attacked with an instrument such as a development tax, then the city ultimately shrinks, and housing prices rise everywhere. By raising the return to real estate investment, this price escalation is likely to spur redevelopment efforts in central neighborhoods.³⁸ Thus, one byproduct of an attack on sprawl at the urban fringe may be upgrading and redevelopment in decaying central neighborhoods. Although there is no formal analysis of such an effect in the literature, the issue could be analyzed by adapting one of the existing models of urban growth with durable housing.³⁹

Many commentators criticize the process of suburbanization, and its attendant “car culture,” as weakening the nation’s social bonds by spreading residences out in low-density patterns that discourage interaction.⁴⁰ Formally, such commentators appear to be arguing that the city’s average population density is a kind of public good, whose level is chosen incorrectly by the decentralized development process. In other words, since the social gains from an increase in average density are not captured by atomistic housing developers, the equilibrium city is too spread out from society’s point of view.

If this argument is correct, then it would appear that supernormal profits could be earned by building relatively dense, large-scale housing developments that internalize the density externality. Such developments, which would tend to limit the extent of urban expansion, are evidenced in several planned communities whose design follows the principles of the “new urbanism.” If such efforts show long-term success, this may indicate that the social-interaction argument has merit. Otherwise, continued low-density

38. See Rosenthal and Helsley (1994) for an empirical analysis of redevelopment.

39. See Brueckner (2000a) for an overview.

40. See, for example, Schwartz (1980).

development on the urban fringe would suggest that consumers prefer the type of neighborhoods that developers have been building all along.

Relaxation of zoning requirements that limit residential density is a prerequisite for the exercise of consumer sovereignty in this area of urban design. While such regulations may simply ratify the previous low-density preferences of consumers, they may also constrain current choices, causing cities to be more spread out than people would like.

Conclusion

When crafting policies to address sprawl, policymakers must recognize that the potential market failures involved in urban expansion are of secondary importance compared with the powerful, fundamental forces that underlie this expansion. For example, while the failure to fully charge for infrastructure costs may impart a slight upward bias to urban expansion, the bulk of the substantial spatial growth that has occurred across the United States cannot be ascribed to such a cause. Instead, this growth mostly reflects fundamentals such as the nation's growing population and rising affluence. Because of the secondary role of market failure, a draconian attack on urban sprawl is probably not warranted. By greatly restricting urban expansion, such an attack might needlessly limit the consumption of housing space, depressing the standard of living of American consumers. Instead, a more cautious approach, which recognizes the damage done by unwarranted restriction of urban growth, should be adopted.

Such caution is a built-in feature of the development taxes, congestion tolls, and impact fees discussed above, which attack sprawl at its source by correcting specific market failures. Urban growth boundaries, by contrast, can easily yield undesirably draconian outcomes because they are not directly linked to the underlying market failures responsible for sprawl. However, because UGBs simply require an extension of existing zoning powers, local policymakers may find them more convenient to use than taxes or tolls. UGBs may therefore end up as the instrument of choice for attacking urban sprawl. One lesson of the discussion is that policymakers should resist the temptation to impose stringent UGBs, recognizing that a substantial restriction of urban growth is likely to do more harm than good.

Comments

Edwin Mills: My views on urban sprawl differ from Jan Brueckner's but by less than might be imagined. He discusses market failure; I discuss governments' failures to get prices right. The unifying observation is that every market failure reflects a failure of governments to get prices right. The difference is in recommended government actions. Those who emphasize market failure mostly want governments to do more. In most cases, I want government to do less; in the crucial area of road pricing, I want governments to do something different from what they have been doing. On issues related to urban sprawl, I believe that government causes of misallocations are patent and simple.

I start with three simple observations. First, "sprawl" is a pejorative term, meaning excessive suburbanization. It would be better to use the neutral term. Second, suburbanization has taken place in every urban area of the world in which it has been studied for most of the past half century; in this country and in at least a few others, for at least a century. Suburbanization has gone farther here than in most countries. We have plentiful cheap land and, until the sprawl police agitated the population, we used it pretty much in accordance with relative prices. Third, there is no intrinsic relationship between suburbanization and traffic congestion. Suburbanization includes dwellings and businesses. I can easily imagine a scenario in which business and population suburbanization occurred so as to reduce commuting distances dramatically in comparison with a monocentric business location pattern. U.S. suburbanization has not been accompanied by falling commuting distances, and I believe this is a much more important issue than excessive suburbanization.

Brueckner understands that invasion of farmland is a nonissue. Except for two world wars, excess, not deficient agricultural output, has been the U.S.

scenario during most of the twentieth century. By now, about half of farm income comes from federal subsidies, and much of the subsidization presumably gets capitalized in farm land prices.⁴¹ No responsible forecast anticipates agricultural shortages for the foreseeable future.

Open space is a similar nonissue. Federal, state, and local governments can and do buy as much land for parks, forests, and so on as their constituents are willing to pay for at fair market prices. In addition, governments can and do buy land for future open space preferences insofar as the democratic process can represent such preferences. The problem comes because governments use the police power to confiscate ownership rights of landowners to preserve open space. The federal government has historically been the worst offender, but state and local governments are now doing the same thing with “growth boundaries” and related police power actions. Imposing the costs of open space preservation on private land owners motivates governments to preserve excessive amounts of open space.

The government action that most promotes excessive suburbanization is local government land use controls. Both central city and suburban governments impose draconian limits on business and residential density—prohibition of multifamily dwellings, minimum lot size requirements, height limitations, floor-area ration limits, and a panoply of other controls. Such controls patently force excessive decentralization of metropolitan areas. They are imposed pursuant to parochial desires of residents to exclude low-income and minority people, whose interests are not represented at the local level since they are not there. As long as local governments can impose land use controls in the interest of their residents, no actions at any government level, such as growth boundaries, can have effects that will not increase distortions. Although motivations are less clear, local governments also impose density and other controls on commercial real estate.

Undoubtedly, the most distorting action of governments in metropolitan areas is the underpricing of transportation. The optimum user fee of an uncongested urban road is the opportunity cost of the land plus the depreciation of the right-of-way plus the operating cost of the road (traffic control, snow removal, and so on) all converted to a vehicle mile basis. I have calculated that a fuel tax of about ten times the typical U.S. level, say \$2.50 per U. S. gallon instead of \$0.25, would be a good approximation to an optimal user fee.⁴² The result would be a gasoline price of about \$4.00 to \$4.50 per gallon, typical of

41. U.S. Department of Agriculture.

42. Mills (1998, pp. 72–83).

prices in Europe and East Asia. Whether my calculation is accurate or not, it is patent that U.S. metropolitan road use is vastly underpriced.

Central cities are not innocent. The city of Chicago imposes severe density limits on downtown office buildings and, in the lakefront communities stretching from the Chicago River north for about five miles (where most of the city's residents of above median income live), has downsized residential densities about 50 percent in the past thirty years. Minorities are proportionately represented on the city council, but typically the councilman has almost complete control over zoning in his or her district. Residents lobby to zone out additional high-density structures. Since the reasons they give do not stand up to the simplest analysis, the central motivation, never mentioned, is presumably the most obvious one: to limit competing housing supply in order to raise the prices of their units.⁴³ Appropriate pricing and inexpensive improvements—traffic control, especially in central cities, and a modest amount of suburban road building—would practically eliminate urban congestion. Car owners would curtail frivolous drives, move closer to workplaces (especially by suburban house swaps that would reduce suburb-to-suburb cross-hauling) and gradually drive smaller, more fuel-efficient vehicles. Residential densities near central business districts and suburban employment subcenters would increase to the extent permitted by land use controls. Fuel tax revenues could be used to reduce distorting real estate taxes, which would also increase business and residential densities if permitted.

Fixed rail commuting is much more underpriced than auto use, but its excess capacity may justify large subsidies, in contrast with road use, which is close to capacity in many metropolitan areas.

The current antisprawl fad is “growth boundaries,” a draconian police power control that Brueckner discusses thoroughly. If one wants to know the effects of growth boundaries, or of less stringent controls of land use conversion from rural to urban uses, plenty of examples are available.⁴⁴ U.S. house prices are less than three times the annual incomes of occupants. In northern Europe and East and South Asia, the ratios are three to ten in many countries. For example, in 1990 house prices were ten times occupants' incomes in Seoul, which had rigidly enforced greenbelt and other controls on land use conversion. At about that time, the government relaxed controls and house prices fell to five times incomes just before the East Asian crisis. In countries with rigid conversion controls, population densities are high, and road con-

43. See Mills (2000).

44. Angel and Mayo (1996).

gestion is at levels unheard of in this country. Even in Canada, which has nearly unlimited amounts of usable land, house prices relative to incomes in Vancouver and Toronto are 50 percent greater than they are on the U.S. side of the border, because of stringent conversion controls.

Nobody gains from pervasive artificial increases in house prices. The one-time housing capital gain is precisely offset by proportionate increases in actual or implicit rents. Elderly people who are about to downsize anyway will gain if they want to conserve part of their capital gain and do not mind leaving an inadequate legacy for housing to their heirs. Of course, if southern California has stringent controls and Phoenix does not, Californians will cash in their houses and move to Phoenix, living handsomely on modest pension or other assets and on the difference between housing costs in Phoenix and California.

Michael Kremer: This paper contains both a positive and normative analysis of urban sprawl. On the positive side, the paper concludes that urban spatial growth is mainly because of increased U.S. population, rising income, and falling commuting costs. On the normative side, it discusses three reasons why growth may be excessive: failure to account for benefits of open space, extra congestion on roads, and failure to charge for the true costs of infrastructure.

Effective measures against urban sprawl will cause cities to be more densely populated. If these measures are appropriate, then the loss in housing consumption will be offset by other gains. If not, consumers will end up worse off.

The paper suggests that road taxes and user fees for infrastructure, along with a potential new development tax, are appropriate responses to the distortions in urban expansion. It then cautions against urban growth boundaries as a blunt instrument.

I agree with most of the analysis. I have a few minor caveats and a couple of suggestions for taking the analysis a step further.

My main problem is with the analysis of the benefits of open space. As the author notes, the typical urban resident may not get huge benefits from preserving open space outside cities in general. Subsidies for agriculture are likely to outweigh any benefits of open space. Similarly, developers are already likely to have to make various payoffs and concessions in exchange for the right to build, even if these are not explicitly treated as infrastructure fees.

To the extent that people do care about preserving open space, it is probably because they drive past it and would like to see prettier views. This

objective suggests a somewhat more targeted approach: instead of taxes on developing agricultural land outside urban boundaries, perhaps these taxes could be on converting land that is visible from well-traveled roads. More broadly, zoning regulations could control unsightly commercial developments along roads. For example, parking lots could be encouraged to be placed behind stores, rather than between stores and the road. I think we may currently be in an equilibrium in which the only way stores can convince customers that parking is available is by displaying the parking in front of the store. Instead, stores could use signs to direct people to parking in the back.

A case could also be made that while the first appearance of a 7-11 spoils the view, the nth appearance does not, suggesting that a nonlinear tax might be appropriate.

Of course, another option would be to follow the lead of the United Kingdom and change property rights to allow hiking through private property, at least on marked trails. Under this type of regime, the benefits of open space outside cities are much clearer.

If the city has a fixed amenity, such as a harbor, property values should increase. This may constitute an important reason for support for urban growth boundaries.

The fiscal externality associated with suburban growth is likely to be quite large, and I expect would outweigh the other effects modeled in the paper in its quantitative importance. For example, when a high-income person moves from Washington, D.C., to Virginia, he or she creates a substantial negative externality for others in Washington.

This paper examines whether incentives for urban expansion are optimal, but it does not go on to examine whether individual communities will have appropriate incentives to control urban growth. This is particularly important when there are different jurisdictions in the city and in the suburbs. If there are many different suburbs, and people drive through several different suburbs, then incentives to control road congestion and views from roads will not be adequate. Incentives to prevent flight from cities by those seeking to avoid taxes will also be inadequate.

References

- Alonso, William. 1964. *Location and Land Use*. Harvard University Press.
- Altshuler, Alan A., and Jose A. Gomez-Ibanez. 1993. *Regulation for Revenue: The Political Economy of Land Use Exactions*. Brookings.
- Angel, Shlomo, and Stephen Mayo. 1996. "Enabling Policies and Their Effects on Housing Sector Performance: A Global Comparison." Istanbul: Habitat II.
- Blomquist, Glenn C., and John C. Whitehead. 1995. "Existence Value, Contingent Valuation, and Natural Resources Damage Assessment." *Growth and Change* 26 (Fall): 573–89.
- Brefle, William S., Edward R. Morey, and Tymon S. Lodder. 1998. "Using Contingent Valuation to Estimate a Neighborhood's Willingness to Pay to Preserve Undeveloped Land." *Urban Studies* 35 (April): 715–28.
- Brueckner, Jan K. 1987. "The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model." In *Handbook of Regional and Urban Economics*, vol. 2, edited by Edwin S. Mills, 821–45. North Holland.
- Brueckner, Jan K. 1995. "Strategic Control of Growth in a System of Cities." *Journal of Public Economics* 57 (July): 393–416.
- . 1997a. "Infrastructure Financing and Urban Development: The Economics of Impact Fees." *Journal of Public Economics* 66 (December): 383–407.
- . 1997b. "Calculation of School Impact Fees." Unpublished paper. University of Illinois at Urbana-Champaign.
- . 2000a. "Urban Growth Models with Durable Housing: An Overview." In *Economics of Cities*, edited by Jacques-Francois Thisse and Jean-Marie Huriot, 263–89. Cambridge University Press.
- . 2000b. "Urban Sprawl: Diagnosis and Remedies." *International Regional Science Review* 23 (April): 160–71.
- Brueckner, Jan K., and Hyun-A Kim. 2000. "Urban Sprawl and the Property Tax." Unpublished paper. University of Illinois at Urbana-Champaign.
- Brueckner, Jan K., and David Fansler. 1983. "The Economics of Urban Sprawl: Theory and Evidence on the Spatial Sizes of Cities." *Review of Economics and Statistics* 55 (August): 479–82.
- Calfee, John, and Clifford Winston. 1998. "The Value of Automobile Travel Time: Implications for Congestion Policy." *Journal of Public Economics* 69 (July): 83–102.
- Ding, Chengri, Gerrit J. Knaap, and Lewis D. Hopkins. 1999. "Managing Urban Growth with Urban Growth Boundaries: A Theoretical Analysis." *Journal of Urban Economics* 46 (July): 53–68.
- Downs, Anthony. 1999. "Some Realities about Sprawl and Urban Decline." *Housing Policy Debate* 10 (4): 955–74.
- Fischel, William A. 1990. "Do Growth Controls Matter? A Review of the Empirical Evidence on the Effectiveness and Efficiency of Local Government Land Use Reg-

- ulation." Lincoln Institute of Land Policy, appearing in Hearing before the Subcommittee on Policy Research and Insurance of the House Committee on Banking, Finance, and Urban Affairs, 101 Cong. 2 sess. August 2, 74–137. Government Printing Office.
- Grant, Peter. 2000. "The Debate over Sprawl Has Only Just Begun," *Wall Street Journal*, November 8, 2000, p. B14.
- Gurwitt, Rob. 1999. "The State vs. Sprawl." *Governing* (January): 18–23.
- Helsley, Robert W., and William C. Strange. 1995. "Strategic Growth Controls." *Regional Science and Urban Economics* 25 (August): 435–60.
- Kim, Kyung-Hwan. 1993. "Housing Prices, Affordability, and Government Policy in Korea." *Journal of Real Estate Finance and Economics* 6 (January): 55–71.
- Lopez, Rigoberto, Farhed Shah, and Marilyn Attolbello. 1994. "Amenity Benefits and the Optimal Allocation of Land." *Land Economics* 70 (February): 53–62.
- Mieszkowski, Peter, and Edwin S. Mills. 1993. "The Causes of Metropolitan Suburbanization." *Journal of Economic Perspectives* 7: 135–47.
- Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review, Papers and Proceedings* 57 (May): 197–210.
- . 1998. "Excess Commuting in U.S. Metropolitan Areas." In *Network Infrastructure and the Urban Environment*, edited by Lars Lundqvist, Lars Gevan Mattson, and Tschango Kim, 73–83. Springer-Verlag.
- . 1999. "Truly Smart 'Smart' Growth." *Illinois Real Estate Letter*. Summer: 1–7.
- . 2000. "Effects of Downzoning on Chicago's North Lakefront Commercial Areas." Northwestern University. Mimeo.
- Muth, Richard F. 1969. *Cities and Housing*. University of Chicago Press.
- Nivola, Pietro S. 1998. "Fat City: Understanding American Urban Form from a Transatlantic Perspective." *Brookings Review* 16 (4): 17–19.
- Orfield, Myron. 1998. "Conflict or Consensus: Forty Years of Minnesota Metropolitan Politics." *Brookings Review* 16 (4): 31–34.
- Rosen, Harvey S. 1985. "Housing Subsidies: Effects on Housing Decisions, Efficiency, and Equity." In *Handbook of Public Economics*, vol. 1, edited by Alan J. Auerbach and Martin Feldstein, 373–420. North Holland.
- Rosenthal, Stuart S., and Robert W. Helsley. 1994. "Redevelopment and the Urban Land Price Gradient." *Journal of Urban Economics* 35 (March): 182–200.
- Rusk, David. 1998. "The Exploding Metropolis: Why Growth Management Makes Sense." *Brookings Review* 16 (4): 13–15.
- Schwartz, Barry. 1980. "The Suburban Landscape: New Variations on an Old Theme." *Contemporary Society* 9 (September): 640–50.
- Small, Kenneth A. 1992. *Urban Transportation Economics*. Harwood Academic Publishers.

- Son, Jae-Young, and Kyung-Hwan Kim. 1998. "Analysis of Urban Land Shortages: The Case of Korean Cities." *Journal of Urban Economics* 43 (May): 362–84.
- Thurston, Lawrence, and Anthony M. J. Yezer. 1994. "Causality in the Suburbanization of Population and Employment." *Journal of Urban Economics* 35 (January): 105–18.
- Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (October): 416–24.
- U.S. Department of Agriculture. Economic Research Service. Annual issues. *Agricultural Income and Finance Situation and Outlook*.
- Wheaton, William C. 1974. "A Comparative Static Analysis of Urban Spatial Structure." *Journal of Economic Theory* 9 (October): 223–37.
- . 1998. "Land Use and Density in Cities with Congestion." *Journal of Urban Economics* 43 (March): 258–72.